





Complete Genome Sequence of the Predatory Bacterium *Ensifer adhaerens* Casida A

 Laura E. Williams,^a  David A. Baltrus,^{b,c} Sean D. O'Donnell,^a Tara J. Skelly,^d Mark O. Martin^e

Department of Biology, Providence College, Providence, Rhode Island, USA^a; School of Plant Sciences, University of Arizona, Tucson, Arizona, USA^b; School of Animal and Comparative Biomedical Sciences, University of Arizona, Tucson, Arizona, USA^c; Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA^d; Biology Department, University of Puget Sound, Tacoma, Washington, USA^e

ABSTRACT We report here the complete genome sequence of the facultative predatory bacterium *Ensifer adhaerens* strain Casida A. The genome was assembled into three circular contigs, with a main chromosome as well as two large secondary replicons, that totaled 7,267,502 bp with 6,641 predicted open reading frames.

Ensifer adhaerens strain Casida A was first isolated in 1980, when Pennsylvania soil samples were enriched for microbes that could prey upon *Micrococcus luteus* (1), and described in greater detail in 1982 (2). *E. adhaerens* represents an isolate with unique biological capabilities in that it is a facultative predator that can target some Gram-positive strains, such as those belonging to the genus *Micrococcus* (3). To date, little is known about the genetics of predation in *Ensifer* strains or which pathways are required for these strains to track *M. luteus*, but a complete genome sequence provides a first step toward identifying pathways of interest in this unique predatory bacterium.

Genomic DNA for PacBio and Illumina libraries was prepared from populations initiated with single colonies and purified as per Baltrus et al. (4). Genomic DNA was independently extracted for the creation of each library type. PacBio sequencing was carried out at the University of North Carolina (UNC) Genomics Core facility on an RSII instrument over four separate single-molecule real-time (SMRT) cells using P6-C4 chemistry. There were 601,168 total polymerase reads across all SMRT cells. Illumina sequencing was performed as part of one multiplexed lane on an Illumina HiSeq platform with 100-bp paired-end libraries at the University of Arizona Genomics Core facility. There were 7,971,492 total reads in the Illumina library, yielding approximately 100× coverage of the genome.

Assembly of PacBio data was performed in an Amazon EC2 instance of SMRTPortal using Hierarchical Genome Assembly Process (HGAP3) (5) with default parameters and an expected genome size of 7.3 Mb. Subread filtering yielded 124,810 subreads with a mean length of 6,143 and an N_{50} of 8,669. Assembly of these subreads generated three large contigs of 4,071,185 bp (93× coverage, 63% GC content), 1,736,943 bp (pCasidaAA, 63% GC, 77× coverage), and 1,459,374 bp (pCasidaAB, 60% GC, 76× coverage). The total published assembly size was 7,267,502 bp. This number of replicons and nucleotide content is consistent with what is found in genomes of strains closely related to *Ensifer adhaerens*, with one main chromosome and two large secondary replicons. Illumina reads, trimmed for quality, were then used to manually correct the PacBio library. Briefly, Illumina reads were mapped to each large contig from the Hierarchical Genome Assembly Process (HGAP) assembly, and single nucleotide variants were mapped using Geneious version 6.0.5 with default parameters. Variants in the PacBio assembly supported by at least 90% of

Received 30 October 2017 Accepted 1 November 2017 Published 22 November 2017

Citation Williams LE, Baltrus DA, O'Donnell SD, Skelly TJ, Martin MO. 2017. Complete genome sequence of the predatory bacterium *Ensifer adhaerens* Casida A. Genome Announc 5:e01344-17. <https://doi.org/10.1128/genomeA.01344-17>.

Copyright © 2017 Williams et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to David A. Baltrus, baltrus@email.arizona.edu.

Illumina reads (with a minimum read depth of 10×) were then manually corrected for the final assembly.

The genome of *E. adhaerens* strain Casida A was annotated using the NCBI Prokaryotic Genome Annotation Pipeline (6) and is predicted to code a total of 6,653 open reading frames (ORFs). The main chromosome contains 3,849 predicted ORFs, 9 rRNAs, 55 tRNAs, and 45 potential pseudogenes. The next largest replicon contains 1,513 predicted ORFs, 3 rRNAs, 3 tRNAs, and 47 potential pseudogenes. The smallest replicon contains 1,279 predicted ORFs, 3 rRNAs, 4 tRNAs, and 95 potential pseudogenes.

Accession number(s). This whole-genome shotgun project has been deposited at GenBank under the accession no. [CP015880](#) (chromosome), [CP015881](#) (plasmid pCasidaAA), and [CP015882](#) (plasmid pCasidaAB). The versions described in this paper are the first versions, CP015880.1, CP015881.1, and CP015882.1, respectively.

ACKNOWLEDGMENTS

We thank Corbin Jones and the UNC Genome Center, as well as the University of Arizona Genome Core for sequencing capacity and much help along the way.

REFERENCES

1. Casida LE. 1980. Bacterial predators of *Micrococcus luteus* in soil. *Appl Environ Microbiol* 39:1035–1041.
2. Casida LE, Jr. 1982. *Ensifer adhaerens* gen. nov., sp. nov.: a bacterial predator of bacteria in soil. *Int J Syst Evol Microbiol* 32:339–345. <https://doi.org/10.1099/00207713-32-3-339>.
3. Martin MO. 2002. Predatory prokaryotes: an emerging research opportunity. *J Mol Microbiol Biotechnol* 4:467–477.
4. Baltrus DA, Nishimura MT, Romanchuk A, Chang JH, Mukhtar MS, Cherkis K, Roach J, Grant SR, Jones CD, Dangl JL. 2011. Dynamic evolution of pathogenicity revealed by sequencing and comparative genomics of 19 *Pseudomonas syringae* isolates. *PLoS Pathog* 7:e1002132. <https://doi.org/10.1371/journal.ppat.1002132>.
5. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Non-hybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <https://doi.org/10.1038/nmeth.2474>.
6. Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity G, Kodira CD, Kyrpides N, Madupu R, Markowitz V, Tatusova T, Thomson N, White O. 2008. Toward an online repository of standard operating procedures (SOPs) for (meta)genomic annotation. *Omic*s 12:137–141. <https://doi.org/10.1089/omi.2008.0017>.